The OECD (Q)SAR Assessment Framework

EU LIFE CONCERT REACH final workshop

Milan, 19 June 2023

Andrea Gissi, PhD
Scientific Officer
European Chemicals Agency

The views expressed in this presentation are those of the author and do not necessarily reflect the official position of the European Chemicals Agency

Table of content

The European Chemicals Agency (ECHA)

About us

We protect humans and the environment by taking action on harmful chemicals

OUR MISSION

We work for the safe use of chemicals

OUR VISION

To be the centre of knowledge on the sustainable management of chemicals for the benefit of citizens and the environment

We implement EU chemicals laws

→ REACH - registration of chemicals

→ Classification, labelling and packaging

→ Biocides

→ PIC – import and export

Our other tasks under EU laws

- → Chemicals in products
- → Poison centres
- → Nanomaterials
- → Persistent organic pollutants
- → Drinking water
- → Exposure limits for workers

Fresh from Press

→ Current status of REACH database + newly registered substances

→ A discussion "***Towards an animal testing-free regulatory system for industrial chemicals***"

- ECHA's activities to promote NAMs
- towards a full replacement of animal testing

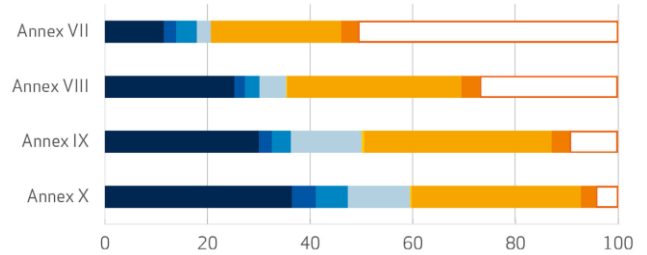


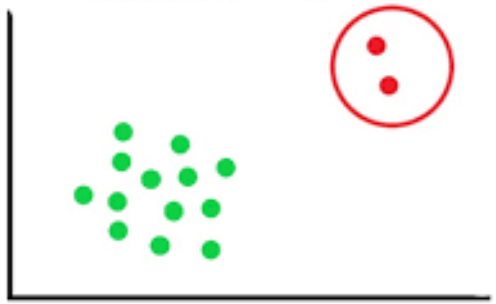
FIGURE 2: Options used to fulfil the information requirements (breakdown per REACH Annex)

The OECD (Q)SAR Assessment Framework

Overview of the project

Valid (Q)SAR model ≠ Valid (Q)SAR result

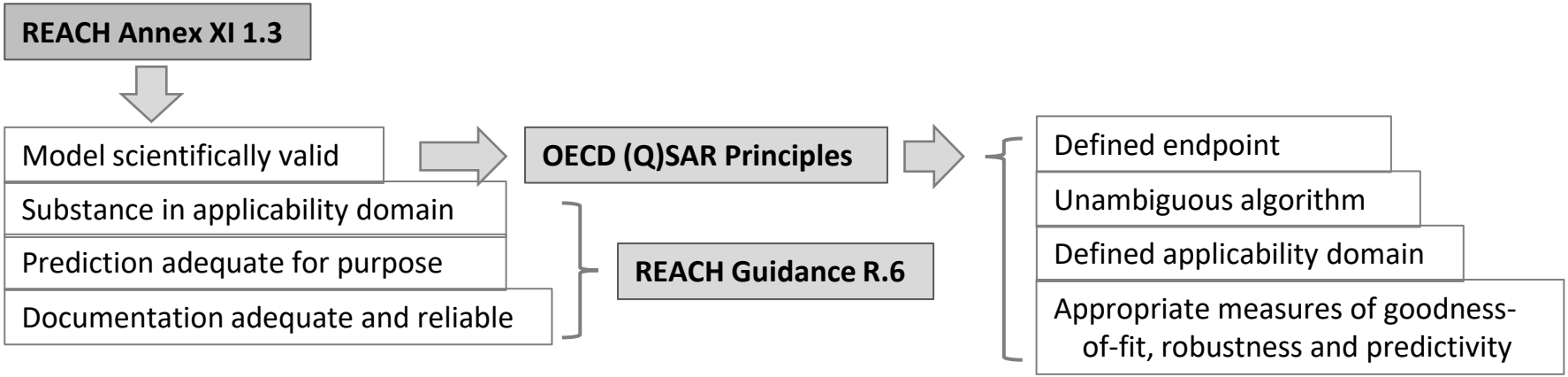
- → The use of (Q)SARs is allowed in many chemical regulations
- → OECD (Q)SAR principles from 2004 cover the scientific validity of **(Q)SAR models**
- → The use of a valid (Q)SAR model does not guarantee the validity of each of its results
- → Need to establish **principles to assess individual results** and a systematic and harmonised **assessment framework** for (Q)SAR models and predictions



The example of REACH

→ Under REACH, (Q)SARs can be used as adaptations to standard information requirements

→ Four conditions to use QSAR results
- scientific validity of the model + three more



OECD principles: OECD ENV/JM/MONO(2007)2: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2

Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals: https://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf

(Q)SAR Assessment Framework: project overview

- **Expert Group** - More than 40 experts from Australia, Canada, Denmark, ECHA, EFSA, Estonia, France, Germany, ICAPO, Italy, Japan, JRC, Netherlands, Sweden, UK, US, Norway
- **Co-leadership** - Italian National Institute of Heath (ISS) and ECHA, Coordinator: OECD
- **Duration**: 24 months

January 2021

Approval of the Project Proposal by WPHA

March 2021

QAF Kick-off meeting

2021-2023

Meetings and drafting (two sub-task groups, predictions and models)

March 2023

End of the project, documents sent to WPHA

28 June 2023

Documents tabled at WPHA for approval

Today (19 June)

September 2023

Publication (if approved in June)

WPHA: OECD Working Party on Hazard Assessment

(Q)SAR Assessment Framework: objectives

- To develop a systematic and harmonised assessment framework for (Q)SAR model predictions
- To revise the QSAR Model Reporting Format (QMRF) and QSAR Prediction Reporting Format (QPRF)
- To address the **uncertainty/confidence** in (Q)SAR predictions
- Applicable irrespective of the modelling technique, the endpoint and the intended regulatory application
- Primarily for regulatory assessors, beneficial for (Q)SAR model developers and users too

Deliverables – QAF Guidance

Two documents:

1. **QAF Guidance**: Text document establishing principles for the assessment of QSAR results and explaining how to assess models and their results

OECD

Organisation for Economic Co-operation and Development

ENV/CBC/HA(2023)4

For Official Use — English - Or. English

12 May 2023

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Working Party on Hazard Assessment

(Q)SAR Assessment Framework

Guidance for the regulatory assessment of (Quantitative) Structure–Activity Relationship models, predictions, and results based on multiple predictions

7th Meeting of the Working Party on Hazard Assessment

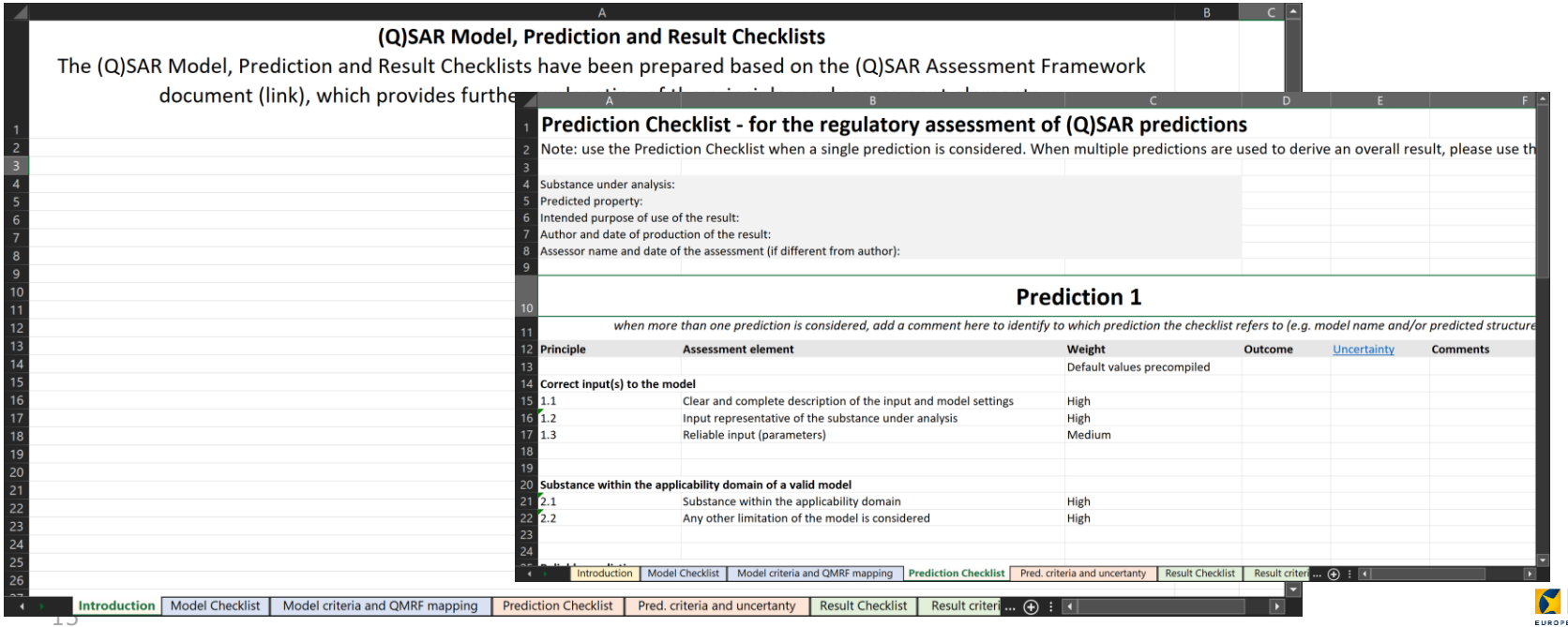
Table of Content

ECHA EUROPEAN CHEMICALS AGENCY

Deliverables – QAF Checklist

Two documents:

2. **QAF Checklist**: Excel document to perform the assessment in practice. Includes the Model Checklist, Prediction Checklist, Result Checklist + examples and explanations



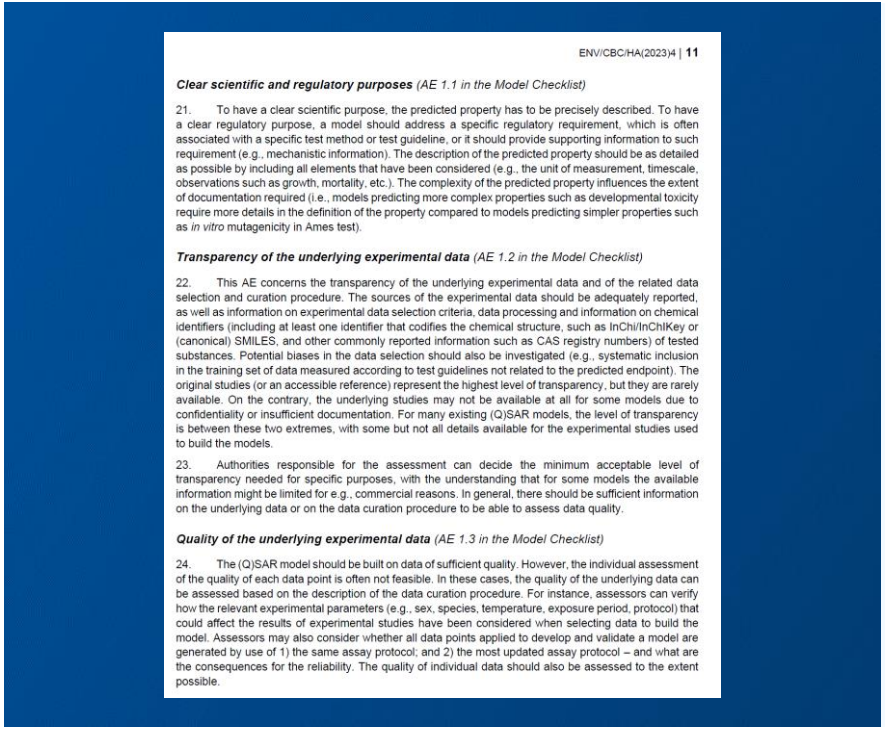
Assessment of (Q)SAR models

Principles for the assessment of (Q)SAR models

- The QAF group agreed that the OECD principles for evaluating the scientific validity of (Q)SAR models remain relevant:

1. Defined endpoint
2. Unambiguous algorithm
3. Defined domain of applicability
4. Appropriate measures of goodness-of-fit, robustness and predictivity
5. Mechanistic interpretation, if possible

QAF Guidance for the assessment of models



ENV/CBC/HA(2023)4 | **11**

Clear scientific and regulatory purposes *(AE 1.1 in the Model Checklist)*

21. To have a clear scientific purpose, the predicted property has to be precisely described. To have a clear regulatory purpose, a model should address a specific regulatory requirement, which is often associated with a specific test method or test guideline, or it should provide supporting information to such requirement (e.g., mechanistic information). The description of the predicted property should be as detailed as possible by including all elements that have been considered (e.g., the unit of measurement, timescale, observations such as growth, mortality, etc.). The complexity of the predicted property influences the extent of documentation required (i.e., models predicting more complex properties such as developmental toxicity require more details in the definition of the property compared to models predicting simpler properties such as *in vitro* mutagenicity in Ames test).

Transparency of the underlying experimental data *(AE 1.2 in the Model Checklist)*

22. This AE concerns the transparency of the underlying experimental data and of the related data selection and curation procedure. The sources of the experimental data should be adequately reported, as well as information on experimental data selection criteria, data processing and information on chemical identifiers (including at least one identifier that codifies the chemical structure, such as InChi/InChIKey or (canonical) SMILES, and other commonly reported information such as CAS registry numbers) of tested substances. Potential biases in the data selection should also be investigated (e.g., systematic inclusion in the training set of data measured according to test guidelines not related to the predicted endpoint). The original studies (or an accessible reference) represent the highest level of transparency, but they are rarely available. On the contrary, the underlying studies may not be available at all for some models due to confidentiality or insufficient documentation. For many existing (Q)SAR models, the level of transparency is between these two extremes, with some but not all details available for the experimental studies used to build the models.

23. Authorities responsible for the assessment can decide the minimum acceptable level of transparency needed for specific purposes, with the understanding that for some models the available information might be limited for e.g., commercial reasons. In general, there should be sufficient information on the underlying data or on the data curation procedure to be able to assess data quality.

Quality of the underlying experimental data *(AE 1.3 in the Model Checklist)*

24. The (Q)SAR model should be built on data of sufficient quality. However, the individual assessment of the quality of each data point is often not feasible. In these cases, the quality of the underlying data can be assessed based on the description of the data curation procedure. For instance, assessors can verify how the relevant experimental parameters (e.g., sex, species, temperature, exposure period, protocol) that could affect the results of experimental studies have been considered when selecting data to build the model. Assessors may also consider whether all data points applied to develop and validate a model are generated by use of 1) the same assay protocol; and 2) the most updated assay protocol – and what are the consequences for the reliability. The quality of individual data should also be assessed to the extent possible.

→ Each principle is broken down to assessment elements (AEs)

→ The Guidance gives more details for each AEs

→Ideally, an acceptable model should fulfil all AEs. However, depending on the purpose of use, evaluators may accept models where not all AEs are fulfilled

Figure: Guidance text with explanation of the AEs for assessing QSAR Models Principle 1: a defined endpoint

Model Checklist

Model 1			
when more than one model is considered, add a comment here to identify to which model the checklist refers to (e.g. model name)			
Principle	**Assessment element**	**Outcome**	**Comments**
Defined endpoint			
1.1	Clear scientific and regulatory purpose		
1.2	Transparency of the underlying experimental data		
1.3	Quality of the underlying experimental data		
Unambiguous algorithm			
2.1	Description of the algorithm and/or software		
2.2	Inputs and other options		
2.3	Model accessibility		
Defined domain of applicability			
3.1	Clear definition of the applicability domain and limitations of the model		
Appropriate measures of goodness-of-fit, robustness and predictivity			
4.1	Goodness-of-fit, robustness		
4.2	Predictivity		
Mechanistic interpretation			
5.1	Plausibility of the mechanistic interpretation		
Conclusion on the model		The conclusion is based on the outcome of the assessment elements as decided by individual authorities	
Comments			

Outcome (for each AE):

- Fulfilled
- Not fulfilled
- Not applicable/assessed, or
- Not documented

Conclusion (for the whole model):

- The model is acceptable for the intended purpose
- The model is not acceptable for the intended purpose
- Documentation insufficient to decide on the acceptance of the model for the intended purpose

Model criteria and QMRF mapping

- A separate spreadsheet of the Checklist provides details, practical advice, examples and mapping to the QMRF for each AE

Checklist for the regulatory assessment of (Q)SAR models
Details on the assessment elements

Principle	Assessment element	Objective	What to check and how	Practical advice	Examples	Mapping to the most relevant QMRF field(s)
Defined endpoint						
1.1	Clear scientific and regulatory purpose	The predicted endpoint is clearly defined in relation to a scientific and/or regulatory purpose.	The predicted endpoint is clearly defined andis consistent with the data used to build the model. For a clear scientific purpose: the predicted endpoint refers to physicochemical, biological or environmental effects o that can be measured and therefore modelled. For a clear regulatory purpose: the predicted endpoint refers to a specific regulatory requirement or test method or test guideline.	The description of the predicted endpoint should be as detailed as possible by including all elements that have been taken into account (e.g. the unit of measurement, timescale, observations such as growth, mortality, etc.).	Clear scientific (and regulatory) purpose: predicted endpoint = "Fish-short term toxicity (96 hours) as LC50 according to the OECD Test Guideline 203". Clear regulatory purpose: Predicted endpoint = "Classification for skin sensitisation according to GHS criteria".	3.2 Endpoint 3.3 Comment on endpoint 3.5. Dependent variable 3.6. Experimental protocol
1.2	Transparency of the underlying experimental data	The documentation is sufficient to independently assess the quality of the experimental data used to build the model for the next assessment element.	Check to what extent the following information is available : - Clear identification of the substances tested (name, structures, SMILES numerical identifiers, etc.); -A (primary) reference to the original studies; - Description of relevant experimental conditions that could affect the prediction (e.g. sex, species, temperature, exposure period, protocol, measurements units); -The original value in the case of data processing before modelling, information on data processing, unit or scale conversion; - Availability of the description of the data aggregation procedure and individual values for datasets where multiple data for the same substance are aggregated for modelling; - Information in the experimental data selection and curation procedure.	It is rare to have full details on each data point used to build the model, but a general description about the experimental data selection and curation procedure can be expected.	Example 1: The model documentation includes the list of substances part of the training set, the experimental values for the predicted property and details or reference for each data point. This assessment element is fulfilled. Example 2: The predicted endpoint is "Bacterial mutagenicity according to OECD TG 471", but the information on the underlying data does include information on the strains tested or presence of metabolic activation. This assessment element is not fulfilled.	3.1 Species 3.4 Endpoints units 3.5 Dependent variable 3.6 Experimental protocol 6.2 Available information for the training set 6.3 Data for each descriptor variable for the training set 6.4 Data for the dependent variable for the training set 6.5 Other information about the training set
1.3	Quality of the underlying experimental data	Ensure that the model is built on data of sufficient quality to obtain acceptable predictions.	- Assess the experimental data curation procedure; - Assess the quality of the data point individually, if possible;	Ideally data points should be evaluated individually. However, especially for large training sets, this may be not possible. In these cases, assessors can verify how the relevant experimental conditions that could affect the results of experimental studies (e.g., sex, species, temperature, exposure period, protocol) have been considered when selecting data to build the model. For models with large training sets, spot check some data points. In some cases, lower data quality can be compensated by large number of data points fitting the same trend.	The model documentation indicates that the predicted endpoint is fish long-term toxicity. The assessment of the data used to build the model shows that the duration of the exposure was not taken into account when selecting data to build the model. It is suspected that some of the data used to build the model refer to results from fish short-term toxicity studies. Outcome: This assessment element is not fulfilled and the model not considered valid for predicting fish long-term toxicity.	3.7 Endpoint data quality and variability 6.6 Pre-processing of data before modelling
Unambiguous algorithm						
2.1	Description of the algorithm and/or software	Ensure that it is clear how the prediction is obtained and that it can be reproduced by others	- Check if a sufficient description of all descriptors and of approach used for their selection and calculation is provided; - Check the availability of a transparent description of the algorithm and/or software, explaining how the predictions were produced. - For fragment/alert based models, the list of the fragments (active, inactive, masks, etc. as relevant) together with information of all substructures and identification of its substituents should be provided. - For equation based models, a description of the equation and all data/descriptors and approach used for their selection should be provided.	An exact description of the algorithm might not be publicly available for commercial models. In such cases, any available relevant information should still be assessed. When the model is implemented in a computer program that is accessible to the assessor, the reproducibility of the results should be possible even for cases when the description of the algorithm is not fully disclosed, and assessors may decide that this is acceptable for some regulatory uses.	User manuals, publications, help files, such as EPISuite help file	4.1 Type of model 4.2 Explicit algorithm 4.3 Descriptors in the model 4.4 Descriptor selection 4.5 Algorithm and descriptor generation 4.6 Software name and version for descriptor generation 4.7 Chemicals/Descriptors ratio 6.1 Availability of the training set
2.2	Inputs and other options	Allowed input formats, pre-processing procedure for the input structures and customisable options/settings are explained.	- Availability of instructions to prepare the input. - Availability of information on the editable options/settings (if any).	The extent of this description depends on the complexity of the computer program. Simple programs with no customisable options require less explanations than programs that allow editing of the settings of the algorithm.	Instructions on the preparation of the input may include instructions how to pre-process salts and tautomers.	1.3 Software coding the model 2.8 Availability of information about the model 6.6 Pre-processing of data before modelling
2.3	Model accessibility	Assess if the model or computer program is or can be available to the assessor.	- Availability of the same model and version described in the documentation	When a different model version is available to the assessor, consider using it and compare the results.	"In vitro mutagenicity (Ames test) alerts" fragment-based model implemented in Toxtree 3.1.0 software available at https://toxtree.sourceforge.net/ has been used for generate a prediction.	1.3 Software coding the model 2.5 Model developer(s) and contact details 2.6 Date of model development and/or publication 2.7 Reference(s) to main scientific papers and/or software package 2.8 Availability of information about the model

Assessment of predictions and results based on multiple predictions

(Q)SAR prediction: an individual output (i.e., the predicted value of a property) of a (Q)SAR model. It can be a continuous or a categorical (two or more categories) output.

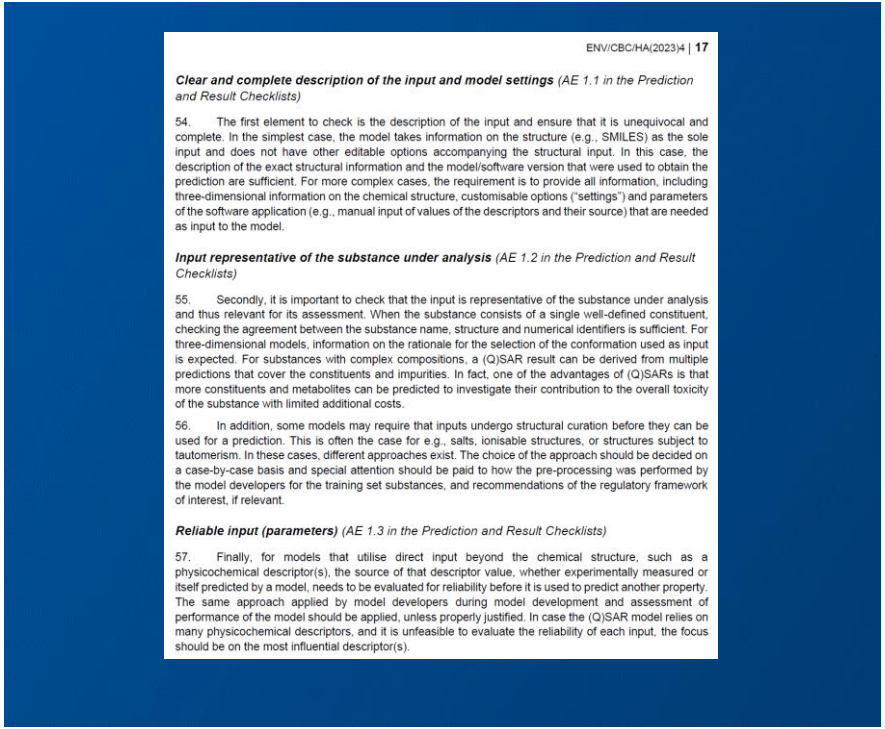
(Q)SAR result: the assessment of a property of a substance based on multiple (Q)SAR predictions.

Principles for the assessment of (Q)SAR predictions

- Four new OECD principles for evaluating (Q)SAR predictions and results based on multiple predictions:
  1. **Correct input** - complete and representative of the substance being analysed, uses reliable parameters
  2. **Substance within applicability domain** – assessment limited to the domain as defined by model developers
  3. **Reliable prediction** – to cover elements that may not be part of the developers' definition of applicability domain
  4. **Outcome fit for purpose** - the usefulness of the computational prediction to answer a specific regulatory question

- For a result based on multiple predictions, first each prediction is assessed individually, and then an additional evaluation step is dedicated to the final result (as explained later)

Guidance for the assessment of (Q)SAR predictions



Clear and complete description of the input and model settings *(AE 1.1 in the Prediction and Result Checklists)*

54. The first element to check is the description of the input and ensure that it is unequivocal and complete. In the simplest case, the model takes information on the structure (e.g., SMILES) as the sole input and does not have other editable options accompanying the structural input. In this case, the description of the exact structural information and the model/software version that were used to obtain the prediction are sufficient. For more complex cases, the requirement is to provide all information, including three-dimensional information on the chemical structure, customisable options ("settings") and parameters of the software application (e.g., manual input of values of the descriptors and their source) that are needed as input to the model.

Input representative of the substance under analysis *(AE 1.2 in the Prediction and Result Checklists)*

55. Secondly, it is important to check that the input is representative of the substance under analysis and thus relevant for its assessment. When the substance consists of a single well-defined constituent, checking the agreement between the substance name, structure and numerical identifiers is sufficient. For three-dimensional models, information on the rationale for the selection of the conformation used as input is expected. For substances with complex compositions, a (Q)SAR result can be derived from multiple predictions that cover the constituents and impurities. In fact, one of the advantages of (Q)SARs is that more constituents and metabolites can be predicted to investigate their contribution to the overall toxicity of the substance with limited additional costs.

56. In addition, some models may require that inputs undergo structural curation before they can be used for a prediction. This is often the case for e.g., salts, ionisable structures, or structures subject to tautomerism. In these cases, different approaches exist. The choice of the approach should be decided on a case-by-case basis and special attention should be paid to how the pre-processing was performed by the model developers for the training set substances, and recommendations of the regulatory framework of interest, if relevant.

Reliable input (parameters) *(AE 1.3 in the Prediction and Result Checklists)*

57. Finally, for models that utilise direct input beyond the chemical structure, such as a physicochemical descriptor(s), the source of that descriptor value, whether experimentally measured or itself predicted by a model, needs to be evaluated for reliability before it is used to predict another property. The same approach applied by model developers during model development and assessment of performance of the model should be applied, unless properly justified. In case the (Q)SAR model relies on many physicochemical descriptors, and it is unfeasible to evaluate the reliability of each input, the focus should be on the most influential descriptor(s).

- Each principle is broken down to assessment elements (AEs)
- Each AE has its own weight and uncertainty
- **Weight:** how important is the AE in the context of use of the prediction (low, medium, or high)
- **Uncertainty**: how confident is the assessor with the outcome

At the end, the overall uncertainty of the prediction is assigned based on the highest uncertainty of high weight AEs.

Figure: Guidance text with explanation of the AEs for assessing QSAR Predictions Principle 1: a correct input

Prediction Checklist

Prediction 1

when more than one prediction is considered, add a comment here to identify to which prediciton the checklist refers to (e.g. model name and/or predicted structure)

Principle	Assessment element	Weight	Outcome	Uncertainty Comments
		Default values		Only for elements that are fulfilled
Correct input(s) to the model				
1.1	Clear and complete description of the input and model settings	High		
1.2*	Input representative of the substance under analysis	High		
1.3	Reliable input (parameters)	Medium		
Substance within the applicability domain of a valid model				
2.1	Substance within the applicability domain	High		
2.2	Any other limitation of the model is considered	High		
Reliable prediction				
3.1	Reproducibility	High		
3.2	Overall performance of the model	Medium		
3.3	Relationship of the substance with the physicochemical, structural and response spaces of the training set of the model	Medium		
3.4	Performance of the model for similar substances	High		
3.5*	Mechanistic and/or metabolic considerations	High		
3.6*	Consistency of information	High		
Outcome is fit for the regulatory purpose				
4.1*	Compliance with additional requirements	High		
4.2*	Correspondence between predicted property and property required by the regulation	High		
4.3*	Decidability within the specific framework	High		
Conclusion on the individual prediciton				
Uncertainty				
Outcome of the assessment (individual prediction)				
Comments				

For each assessment element (AE):

→ **Weight**

- Low; Medium; High

→ **Outcome:**

- Fulfilled; Not fulfilled; Not applicable/assessed; Not documented

→ **Uncertainty:**

- Low; Medium; High

By default, high uncertainty to AEs that are not fulfilled or not documented

Prediction Checklist

Prediction 1				
when more than one prediction is considered, add a comment here to identify to which prediciton the checklist refers to (e.g. model name and/or predicted structure)				
Principle	**Assessment element**	**Weight**	**Outcome**	**Uncertainty Comments**
		Default values		Only for elements that are fulfilled
Correct input(s) to the model				
1.1	Clear and complete description of the input and model settings	High		
1.2*	Input representative of the substance under analysis	High		
1.3	Reliable input (parameters)	Medium		
Substance within the applicability domain of a valid model				
2.1	Substance within the applicability domain	High		
2.2	Any other limitation of the model is considered	High		
Reliable prediction				
3.1	Reproducibility	High		
3.2	Overall performance of the model	Medium		
3.3	Relationship of the substance with the physicochemical, structural and response spaces of the training set of the model	Medium		
3.4	Performance of the model for similar substances	High		
3.5*	Mechanistic and/or metabolic considerations	High		
3.6*	Consistency of information	High		
Outcome is fit for the regulatory purpose				
4.1*	Compliance with additional requirements	High		
4.2*	Correspondence between predicted property and property required by the regulation	High		
4.3*	Decidability within the specific framework	High		
Conclusion on the individual prediciton				
Uncertainty				
Outcome of the assessment (individual prediction)				
Comments				

Conclusion

→ **Uncertainty of the prediction**

- Low; medium; High

Based on the highest uncertainty of high weight AEs.

→ **Outcome of the assessment**

- Acceptable for the intended purpose;
- Not acceptable for the intended purpose;
- Documentation insufficient to decide on the acceptance for the intended purpose.

The document suggests to accept predictions with low or medium uncertainty

Predictions criteria and QPRF mapping

→ Also for predictions and results, a separate spreadsheet of the Checklist provides details, practical advice, examples and mapping to the QPRF for each AE

→ In addition, there is a section dedicated to how to assign the uncertainty level

Principle	Practical advice	Examples	Uncertainty		Mapping to most
			This table offers guidance on how to assign the uncertainty level of each assessment element. *To assign the uncertainty for elements that are fulfilled, refer to the explanation in the column.* *For elements that are not fulfilled or not documented, high uncertainty should be assigned by default unless a valid justification is provided.* *For elements that are not applicable/assessed, leave empty* *NOTE: some examples include numeric values to explain more concretely how to proceed with the assessment . However, acceptable values depend on the predicted property and purpose of use of the prediction. The values used as examples should not be intended as thresholds established by the project.*		
Correct input(s) to			**Exaplanation of the uncertainty level**	**Examples**	
1.1	If the input is incomplete but the assessors are still able to reproduce the prediction, then the weight of this element in the overall assessment is lower.	Example 1: in case the model accepts as input the structure in form of SMILES, it is not sufficient to indicate as input the substance name and/or its numerical identifiers (such as CAS or EC numbers). Names and numerical identifiers may not unequivocally identify the SMILES that has been used as input. The exact SMILES used as input needs to be specified. Example 2: in case the model accepts as input three-dimensional structures, it is not sufficient to indicate as input the SMILES of the structure. Information on the three-dimensional structure, such a .mol file or equivalent, is needed.	**Low:** input structure(s) and model settings are fully described **Medium:** some minor aspects of the input structure(s) and model settings are not clearly described **High:** some important aspects of the input structure(s) and model settings are not clearly described	A model requires SMILES and optionally logKow as input to generate a prediction. Low: SMILES and logKow provided Medium: SMILES provided, logKow not provided High: only CAS number provided, but CAS/SMILES association is ambiguous. **NOTE:** the reliability of logKow is assessed under AE 1.3	5 Input (all fields
1.2	The comparison can be done using expert judgment or by using publicly available information and tools that associate structures with names or other identifiers. If the model distinguishes the different tautomeric forms and generates different predictions, then it is important to indicate which form was used as input and justify the selection. If different tautomeric forms are investigated and produce the same prediction, this should also be indicated. If the model documentation indicates how to pre-process the input structure, possibly including how to represent tautomeric groups, these indications should be followed. Alternatively, the user should (if possible) use as input the structure in the tautomeric form that would be predominant if the corresponding experimental test were performed to measure the property of interest. Another option is to predict different forms and to calculate either a reasonable worst-case or an average, eventually weighted according to the abundance of the different forms.	Example 1: the substance under analysis is "formhaldeyde". The SMILES "C=O" is used as input. Using available resources, the correspondence between the name and the SMILES is verified. Example 2: the substance under analyis is a salt formed by an inorganic cation and an organic anion. The model does not accept the SMILES that includes both ions. The model documentation indicates that for salts, only the neutralised organic part should be used as input. The assessment consists in checking that the correct pre-processing has been followed. *Example 3 (for multiple predictions): the substance is formed by two major constituents. If two separate predictions are provided for the constituents, then the assessment element is fulfilled	**Low:** the composition of the substance under analysis is well covered by the input structure(s) **Medium:** the composition of the substance under analysis is mostly covered by the input structure(s) **High:** some constituents of the substance under analysis are not covered by the input structure(s)	The prediction refers to a substance that includes three constituents (one major constituent, one minor constituent and one impurity) in its composition. Low: predictions for all three constituents are provided Medim: predictions for two constituents are provided, impurity not considered High: only the prediction for the major constituent is provided	5 Input (all fields 2 Substance (all
1.3	Parameters that are automatically calculated by the model or software do not need to be evaluated at this stage.	An aquatic toxicity prediction is obtained from a model based on logKow. The prediction is generated by using as input an logKow defined by the user. The reliability of the user defined logKow needs to be verified.	**Low:** the values of the additional input parameters are associated with low uncertainty **Medium:** the values of additional input parameters are associated with medium uncertainty **High:** the values of additional input parameters are associated with high uncertainty	A model that requires manual input of logKow is used to generate a prediction. Low: the logKow value used as input is the result of a reliable experimental study Medium: the logKow value used as input is predicted by a QSAR model. No details are provided to assess its reliability. High: the logKow value used as input is predicted by a QSAR model. The prediction is unreliable, but it is the only available estimate.	5.2 Descriptors

(Q)SAR results based on multiple predictions

Cases that consider multiple predictions include:

→ Predictions from different models for the same structure;

→ Predictions from the same models for different structures (such as the multiple constituents of a substance or for the substance under analysis and its metabolites);

→ A combination of the above.

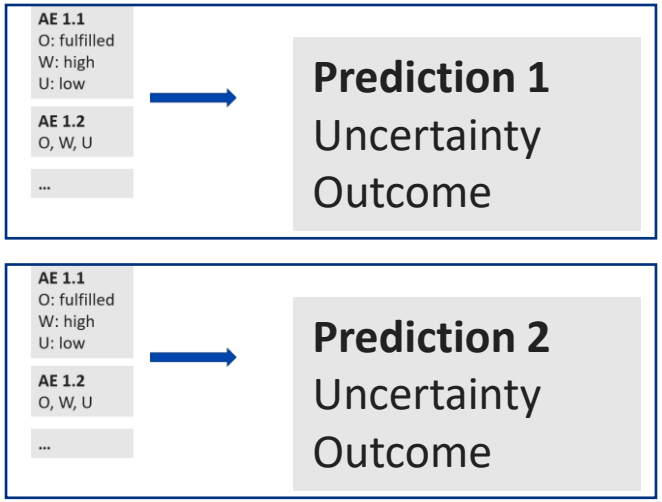
Assessment workflow for results from multiple predictions

1. Within the Result Checklist, complete a checklist for each prediction individually (for complex cases, start by addressing multiple predictions associated with the same structure, and then consider the predictions for different structures)
2. Assess the additional AE:
   - Correct determination of the final result from individual predictions
3. Determine the uncertainty of the final result by weighing the uncertainty of individual predictions (e.g. consistent independent predictions lower uncertainty)
4. Decide on the acceptability of the result (the document suggests to accept results with low or medium uncertainty)

Workflow for assessing results from multiple predictions

Assessment element (AE)
Outcome (O): fulfilled, not fulfilled, not documented, not applicable
Weight (W): low, medium, high
Uncertainty (U): low, medium, high
Conclusion: results acceptable, not acceptable, insufficient documentation

1. Assess predictions individually



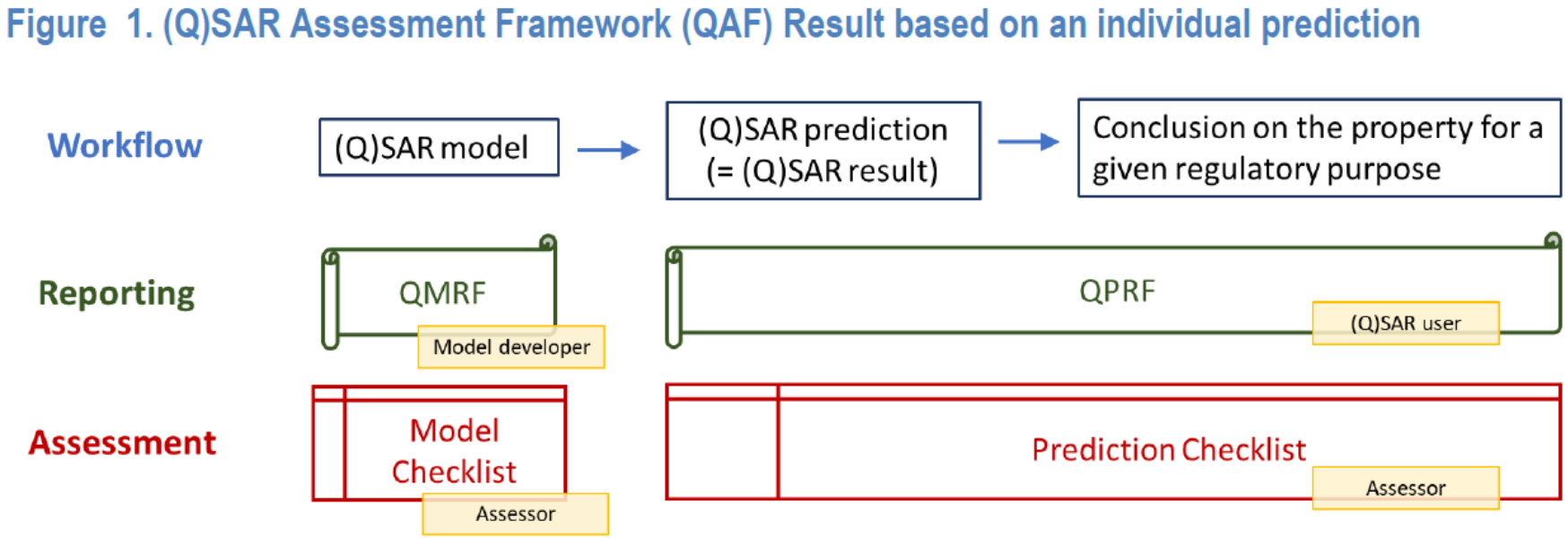
2. Check how the final result is determined (AE 5.1)

(Q)SAR result

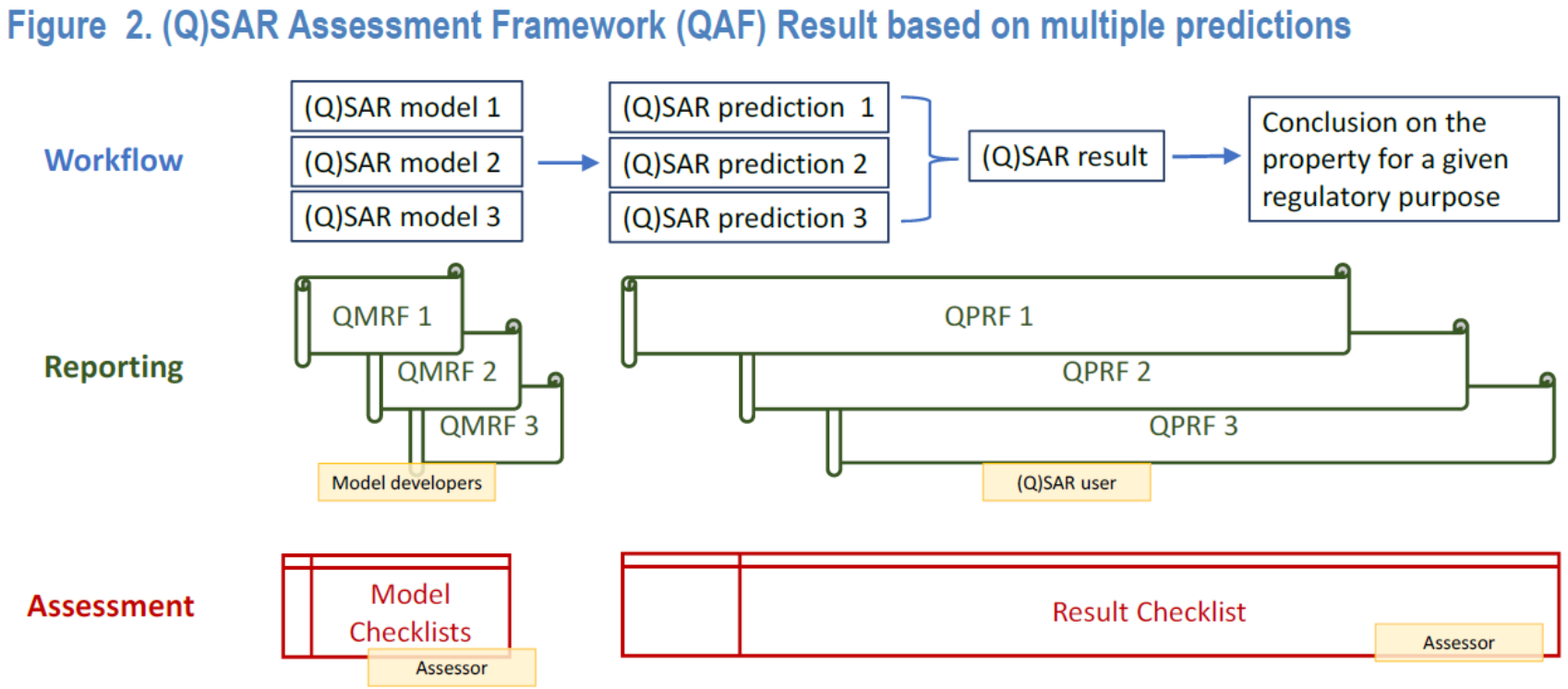
3. Conclusion based on the level of uncertainty and purpose of use

Conclusion on the result
Uncertainty
Outcome

Visual abstract 1/2



Visual abstract 2/2



Conclusions

What is next

→ The OECD QAF expert group identified the following areas for further work:

- **Endpoint specific case studies**
- **Reporting** (revise the QMRF and design a report for results from multiple predictions)
- Technical guidance on **how to validate models** (i.e. measure the performance), especially in terms of external validation

OECD (Q)SAR assessment framework - Take home messages

The publication by OECD of the QAF documents is expected in September 2023

Establishes new OECD principles for the assessment of (Q)SAR predictions and results from multiple predictions, and provides guidance and checklists for their assessment

The QAF will be the reference point for the regulatory assessment of (Q)SARs

With a systematic and harmonised assessment framework, the QAF will benefit regulators first, and then model developers and (Q)SAR users too

Acknowledgments

QAF Contributors

→ **Co-leads from ISS**
- Olga Tcheremenskaia
- Cecilia Bossa
- Chiara Battistelli

→ **OECD secretariat**
- Patience Browne
- Tomoko Aoyagi

→ **QAF Group Members**

If you want to go fast go alone.
If you want to go far go together.
(African Proverb)

QAF Group members*

Jessica Barkas (US Environmental Protection Agency (EPA))
Tara S. BARTON-MACLAREN (Health Canada)
Chiara Laura BATTISTELLI (National Institute of Health (ISS))
Michael BEKING (Environment and Climate Change Canada)
Mark BONNELL (Environment and Climate Change Canada)
Cecilia BOSSA (National Institute of Health (ISS))
Edoardo CARNESECCHI (European Food Safety Authority (EFSA))
Daniel Chang (US Environmental Protection Agency (EPA))
PAULA ECATERINA CHIRA (Ministry of Environment, Waters and Forests)
Anna Cruz (Australian Industrial Chemicals Introduction Scheme (AICIS))
Joop de Knecht (National Institute for Public Health and the Environment (RIVM))
Ayako FURUHAMA (National Institute of Health Sciences (NIHS))
Andrea Gissi (European Chemicals Agency (ECHA))
Marius Gudbrandsen (Climate and Pollution Agency)
Jörgen Henriksson (Swedish Chemicals Agency)
Matthias HERZLER (German Federal Institute for Risk Assessment (BfR))
Rune HJORTH (Danish Environmental Protection Agency)
Tina Hofmaier (Austrian Agency of Health and Food Safety (AGES))
Masashi HORIE (Ministry of Economy, Trade and Industry)
Miriam JACOBS (UK Health Security Agency)
Nobheetha Jayasekara (Australian Industrial Chemicals Introduction Scheme (AICIS))
Shinano KAWAHARA (Ministry of Economy Trade and Industry (METI))
Jeannette Koenig (Federal Institute for Risk Assessment (BfR))
Sunil KULKARNI (Health Canada)
Knud Ladegaard Pedersen (Danish Environment Protection Agency)
Donna Macmillan (Humane Society International (HSI))
Uko Maran (University of Tartu)
Julia Havilland MARKS (Department of Agriculture, Water and the Environment)
Todd Martin (US Environmental Protection Agency (EPA))
Enrico MOMBELLI (Institut National de l'Environnement Industriel et des Risques (INERIS))
Nikolai Georgiev NIKOLOV (National Food Institute)
Audrey Pearson (Environment Agency)
Martin Phillips (U.S. Environmental Protection Agency (EPA))
Prachi PRADEEP (Federal Institute for Risk Assessment (BfR))
Qinghong PU (Department of the Environment)
Sarah Robinson (Australian Industrial Chemicals Introduction Scheme (AICIS))
Emiel Rorije (National Institute of Public Health and the Environment (RIVM))
Yuki SAKURATANI (National Institute of Technology and Evaluation)
Gerrit SCHÜÜRMANN (UFZ Centre for Environmental Research)
Rositsa Serafimova (European Food Safety Authority (EFSA))
Yongkwon SONG (Permanent Delegation of Korea to the OECD)
Olga TCHEREMENSKAIA (Istituto Superiore di Sanita (ISS))
David Tobias (EPA-Office of Pollution Prevention and Toxics)
Manuela Trobej (Austrian Agency of Health and Food Safety (AGES))
Petra Van Kesteren (National Institute for Public Health and the Environment (RIVM))
Eva Bay WEDEBYE (National Food Institute)
Andrew WORTH (European Commission (Joint Research Centre))
Takashi YAMADA (National Institute of Health Sciences (NIHS), Division of Risk Assessment, Center for Biological Safety and Research)

*official list from OECD

Thank you

andrea.gissi@echa.europa.eu

echa.europa.eu/subscribe

Connect with us

echa.europa.eu/podcasts

European Chemicals Agency

@one_healthenv_eu

@EU_ECHA

@EUECHA

EUchemicals